Spring 2011

Sequencing of Amine Nucleoside Polymers: Understanding the Fidelity of DNA-Directed Reductive Amination

Sarah M. Anderson
Bethel University

Sequencing of Amine Nucleoside Polymers: Understanding the Fidelity of DNA-Directed Reductive Amination

Sarah M. Anderson

Dr. Trey Maddox, Advisor

A Research Proposal Submitted for Completion

of the Bethel University Honors Program

Bethel University

May 2011

Introduction

Template directed polymerization occurs when a parent molecule serves as the template for the synthesis of a complementary daughter molecule. A prime example is DNA replication (Figure 1) where one DNA strand serves as the template for a daughter strand. Because template directed polymerization produces copies or complements of an original, it is an efficient mechanism for the transfer of information. However, biological examples of template directed polymerization require highly specific enzymes, such as DNA polymerase, to catalyze the polymerization reactions *in vivo*. This need for enzymes has limited the scope of template directed polymerizations due to the highly specific nature of the substrates accepted by these enzymes. Therefore, if the need for enzymes could be removed, template directed polymerization would be broadened for use in new applications.

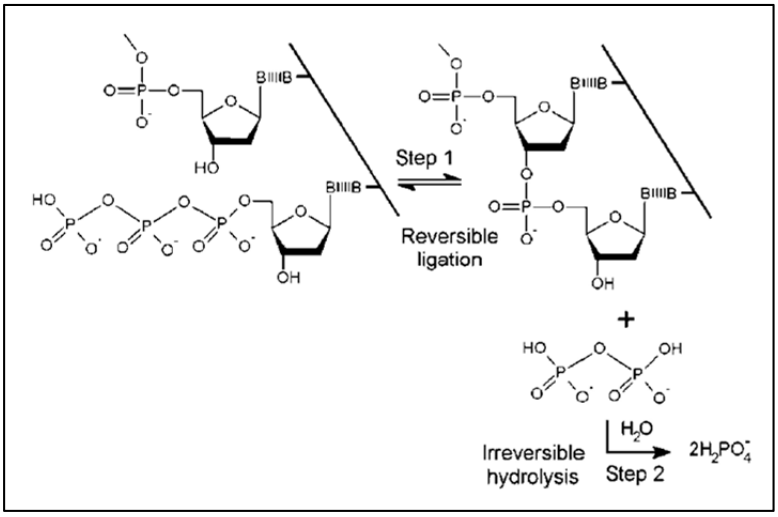

Figure 1. General two-step reaction of DNA replication.

Reductive amination is the two-step reaction between an amine and an aldehyde. These steps, equilibrium formation of the imine followed by irreversible reduction to the amine, mimic those for DNA replication (reversible ligation followed by irreversible hydrolysis).[1] Therefore,

nucleosides modified to undergo reductive amination could potentially participate in non-enzymatic template directed polymerization with a DNA template (Figure 2).

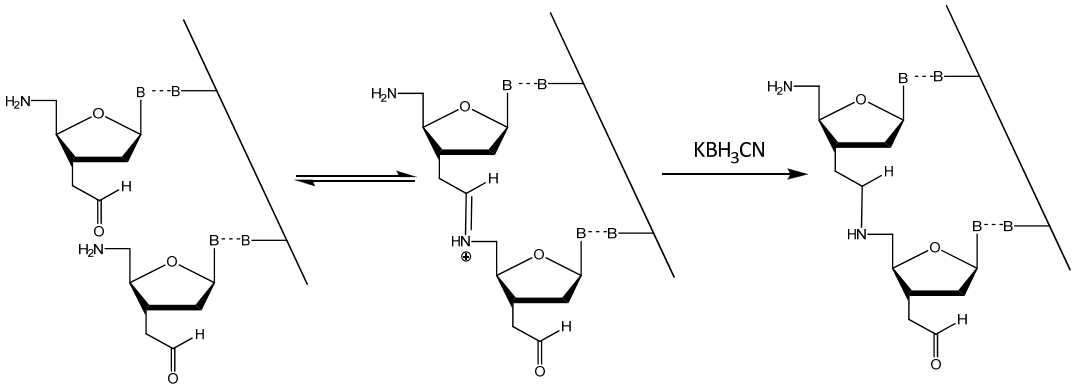

Figure 2. DNA-directed reductive amination.[2]

The modification of the nucleosides requires changing the 3'-hydroxyl to an aldehyde and the 5'hydroxyl to an amine (Figure 3). Polymerizations using these modified monomers have been carried out forming amine nucleoside polymer (ANP) as seen in Figure 2. The results showed that template directed polymerization by reductive amination proceeds with chain length specificity.[3]

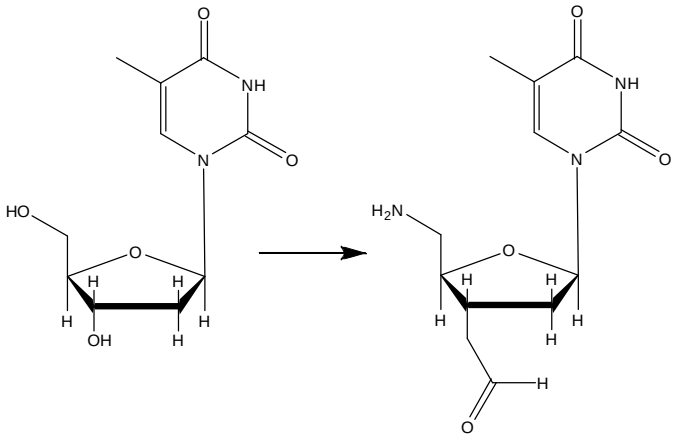

Figure 3. Modification of the thymidine nucleoside.

Additionally, initial experiments have demonstrated that DNA-directed reductive amination proceeds with some degree of fidelity.[3] However, further studies must be performed to better understand the extent of fidelity and how the kinetics of this system can be controlled. These studies will require the ability to sequence the resulting ANP products formed under different reaction conditions. However, traditional sequencing techniques for nucleic acids and polypeptides will not work for sequencing ANP because the techniques only work for polymers with those respective backbones. Therefore, to further study DNA-directed reductive amination, a methodology for sequencing the ANP products must be developed.

Background

To unambiguously determine whether DNA-directed reductive amination proceeds with fidelity, the ANP products must be sequenced and compared to the template. While the more specific sequencing methods for nucleic acids and polypeptides cannot be used to sequence ANP, the more general but equally viable methodology of tandem mass spectrometry could be extended to the sequencing of ANP. Mass spectrometry fragments samples and separates the product fragments by their mass-to-charge ratios. If an ANP was sequentially fragmented along the backbone, the differences in mass between the bases would give the necessary information to construct the sequence.

Tandem mass spectrometry (MS/MS) allows sequencing of large biomolecules by fragmentation. In traditional MS/MS using the product ion scan, the sample is injected into the instrument and ionized in the ion source. The $[M+1]^+$ precursor ion is selected in the first mass analyzer and fragmented in the collision cell, and the fragmented product ions are sorted by their mass-to-charge (m/z) ratios in the second mass analyzer.[4-5]

The ion source is where the sample gets transferred to the gas phase and ionized. Electrospray ionization (ESI) is a "soft" ionization technique that works well for ionizing large biomolecules like polypeptides[4], oligonucleotides[5-6], oligosaccharides[5], polypeptides[8], and polyamines[7] because it efficiently transfers biomolecules to the gas phase with minimal fragmentation.[4,6] In ESI, the sample is injected into the instrument as a liquid, which forms charged droplets. These droplets shrink as the solvent evaporates leaving only the $[M+1]^+$ to travel to the first mass analyzer.[9] Since ANP is of similar scale to the previously mentioned biomolecules, ESI would be the ionization source of choice for sequencing by MS/MS.

The triple quadrupole mass analyzer is the most common tandem-in-space instrumentation; it consists of three quadrupoles.[10] A quadrupole is made up of two pairs of rods that have opposite ac voltages applied to them. Additionally, a dc voltage is scanned causing the *m/z* ratio separation. Only ions with certain *m/z* ratios pass straight through the quadrupole due to the ac and dc voltages; the trajectories of the other ions cause these ions to hit the rods where they convert to neutral molecules.

In the triple quadrupole instrument, the first quadrupole selects the precursor ion by acting as a mass filter. The second quadrupole is the collision chamber. No dc voltage is applied to this quadrupole, which traps the precursor ion in a high concentration of collision gas to induce the formation of product ions. In the third quadrupole, the dc voltage is again scanned allowing mass analysis of the product ions.

To increase the sensitivity of the instrumentation, the ESI/MS/MS could be configured to undergo nozzle-skimmer (NS) collisionally-activated dissociation (CAD), also called in-source collision-induced dissociation (CID).[4-5,8,11] NS CAD is performed by varying the NS voltage bias, which causes collisions of the precursor ions in the source. When using NS CAD, the

MS/MS instrument is run in the precursor ion mode where the first mass analyzer is scanned and the second mass analyzer is held at a constant voltage.[5] Because the initial fragmentations occur in the low pressure region before any filtering or scanning, the $[M+1]^+$ parent ion cannot be preselected.[4] However, NS CAD has the advantage of increasing the product ion resolution and intensity because of the increased fragmentation in the source.[8]

For sequencing to work, fragmentation must only occur along the backbone. This cleavage can be selected for by using low collision energies.[4] Because the collision energy correlates to the extent of fragmentation, a low collision energy causes low energy dissociations, meaning the weakest bonds will cleave first and rearrangements are unlikely. Polyamines have been fragmented by inert gases with collision energies under 20 eV (5 to 20 eV).[7] Collision energies between 40 and 75 eV have been used to sequence oligonucleotides without cleaving the bases from the backbone.[6] Because the collision energy required to cleave amine linkages is less than the that required to cleave phosphodiester linkages, fragmentation of the ANP backbone should occur before cleaving of either the bases or the sugars.

Analyzing mass spectra to determine sequence involves three steps.[6] First, the terminal residues of the ANP must be recognized. Second, the ANP sequence must be constructed independently from each terminus. If any region of the ANP is not sequenced, its mass can be determined by taking the difference between the experimental molecular mass and the sum of the sequenced masses. These sequences are then aligned to find overlap. Third, the proposed sequence must be compared to the experimental molecular mass and rejected if incorrect.

The triple quadrupole ESI/MS/MS has been utilized to successfully sequence polypeptides[4], oligosaccharides[5], oligonucleotides[5-6], and polyamines.[7] NS CAD using the same instrument has been used to successfully sequence oligonucleotides[5], oligosaccharides[5], and

polypeptides.[8,11] Because ANP is a hybrid between an oligonucleotide and a polyamine, a methodology that can sequence both oligonucleotides and polyamines should have the capabilities to sequence the ANP molecules. Therefore, triple quadrupole ESI/MS/MS is a viable option for sequencing the ANP products from DNA-directed reductive amination.

Proposed Research

The ANP products from DNA-directed reductive amination will be sequenced by triple quadrupole ESI/MS/MS in the product ion scan mode, an instrumental setup that has been used to successfully sequence polypeptides[4], oligonucleotides[5-6], oligosaccharides[5], and polyamines.[7]

Tests will be done using nitrogen[7] and argon[4-6] as the collision gases to determine which produces the best backbone fragmentation. Because argon is a more massive gas, it is expected that it will cause more fragmentation than nitrogen will. To induce only backbone fragmentation, the collision energy will be set to 20 eV. If too much fragmentation occurs, the collision energy will be decreased by increments of five to 5 eV.[7] If too little fragmentation occurs, the collision energy will be increased to 40 eV, an energy used to fragment oligonucleotides.[6]

The initial sequencing experiments will be run with homogeneous templates ($T_nT_nT_nT_nT_nT_nT_nT$) to optimize four parameters: the collision gas, the collision energy, the temperature, and the pressure. Ideally, this sequence should produce spectra with evenly spaced peaks. Once the parameters have been optimized, heteropolymers ($A_nT_nA_nT_nA_nT_nA_nT_nA_nT$ and $A_nA_nA_nA_nT_nT_nT_nT$) will be analyzed to test the sequencing capabilities of the mass spectrometric methodology.

If the initial optimization spectra are low resolution, the instrumentation will be changed to run NS CAD in the precursor ion mode.[5] This method should increase the amount of product

ions giving higher resolution spectra. The optimization experiments would need to be redone in this mode.

Once the instrumentation has been optimized, the mass spectrometer will be coupled to liquid chromatography to allow real-time sequencing of the growing ANP. Other MS instrumentations have been done to sequence polyamines and oligonucleotides but cannot be coupled to an LC making ESI/MS/MS the best methodology for sequencing ANP.[12]

If the terminal residues of the ANP cannot be distinguished in the mass spectrum, they could be capped before sequencing. The fluorenylmethyoxycarbonyl (Fmoc) group (223.2 g/mol) is an amine protecting group that can easily be added to a primary amine mixing with sodium carbonate and deionized water in dioxane and stirring at room temperature followed by pouring in a water/brine mixture and extracting with dichloromethane. Because Fmoc does not have the same molecular mass as any of the bases (Table 1), it could be used to cap the 5'-terminal end of the ANP. The 3'-aldehyde could be capped with an acetal. Unfortunately, if it is necessary to cap the terminal ANP, real-time sequencing would not be an option.

Table 1. Molecular masses of the nitrogenous bases.

Base	**Molecular Mass (g/mol)**
Thymidine	126.1
Adenine	135.1
Cytosine	111.1
Guanine	151.1

Acknowledgements

Bethel University Department of Chemistry

References

(1) Goodwin, J. T.; Lynn, D. G. *J. Am. Chem. Soc.* **1992**, *114*, 9197-9198.

(2) Zhan, Z. J.; Lynn, D.G. *J. Am. Chem. Soc.* **1997**, 119, 12420-12421.

(3) Li, X.; Zhan, Z. J.; Knipe, R.; Lynn, D. G. *J. Am. Chem. Soc.* **2002**, *124*, 746-747.

(4) Buré, C.; Lange, C. *Current Organic Chemistry* **2003**, *7*, 1613-1624.

(5) Chen, H.; Tabei, K.; Siegel, M. M. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 846-852.

(6) Ni, J.; Pomerantz, S. C.; Rozenski, J.; Zhang, Y. McCloskey, J. A. *Analytical Chemistry* **1996**, *68*, 1989-1999.

(7) Häkkinen, M. R.; Keinänen, T. A.; Vepsäläinen, J.; Khomutov, A. R.; Alhonen, L.; Jänne, J.; Auriola, S. *J. Pharm. Biomed. Anal.* **2007**, *45*, 625-634.

(8) Huddleston, M. J.; Bean, M. F.; Carr, S. A. *Anal. Chem.* **1993**, *65*, 877-884.

(9) Silverstein, R. M.; Webster, F. X.; Kiemle, P. J. *Spectrometric Identification of Organic Compounds*, 7th ed. John Wiley & Sons: New York, 2005; pp 6-8.

(10) Skoog, D. A.; Holler, F. J.; Crouch, S. R. *Principles of Instrumental Analysis*, 6th ed. Thomson Brooks/Cole: Belmont, 2007; p 575.

(11) Zhai, H.; Han, X.; Breuker, K.; McLafferty, F. W. *Anal. Chem.* **2005**, *77*, 5777-5784.

(12) Kenny, P. T. M.; Goodnow, R. A., Jr.; Konno, K.; Nakanishi, K. *Rapid Commun. in Mass Spectrom.* **1989**, *3*, 295-297.